DOCUMENT RESUME

ED 220 525 TM 820 546

AUTHOR Alvord, Gregory; Macready, George B.
TITLE Comparing Models from Non-Subsuming Parametric Families.
PUB DATE Mar 82
NOTE 17p.; Paper presented at the Annual Meeting of the American Educational Research Association (66th, New York, Ny, March 19-23, 1982).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Comparative Analysis; *Goodness of Fit; Latent Trait Theory; *Mathematical Models; *Probability; *Statistical Analysis; Statistical Studies
IDENTIFIERS Likelihood Ratio Criterion; Rasch Model

ABSTRACT

The Pearson and likelihood ratio statistics are frequently used for assessing the absolute fit of probability models. Researchers are often interested in comparing fits provided by different models which may have a subsuming or non-subsuming relation. A subsuming relation exists when the parameters of the reduced model form a subset of those contained in the subsuming, or full, model. A non-subsuming relation exists when the defining parameters in neither model form a subset of those defining the other model. A general "mixture" probability model which incorporates two non-subsuming probability models and a strategy for assessing fit provided by each of the component models is described. A detailed description of the strategy for selecting a preferred model is outlined. Applications of the suggested procedure are considered for the special case in which the Rasch latent trait model and Latent State Mastery model are compared as to relative preference when data are generated from each model. (DWH)

ED220525

COMPARING MODELS FROM NON-SUBSUMING PARAMETRIC FAMILIES

Gregory Alvord

&

George B. Macready

University of Maryland

Paper presented at the American Educational Research Association 1982 annual meeting, New York.

ABSTRACT

A "mixture" probability model which incorporates two psychometric models from separate parametric families is introduced and the application of this model in selecting a preferred component model is described. Example applications of the suggested procedure are considered for the special case in which the Rasch latent trait model and the Latent State Mastery model are compared as to their relative preference when data are generated from each of these models.

Introduction

Within the realm of probabilistic modeling, researchers are frequently interested in ascertaining how well investigated models account for manifest data. Two frequently used statistics for assessing the absolute fit of probability models are the Pearson and the likelihood ratio statistics. These statistics are defined respectively in equations (1) and (2).

$$\chi^2_P = \sum_{i=1}^{n_r} \frac{(O_i - E_i)^2}{E_i} \qquad (1)$$

and

$$\chi^2_{LR} = -2 \sum_{i=1}^{n_r} O_i \ln\left(\frac{E_i}{O_i}\right) \qquad (2)$$

where

O_i = observed frequency for the i^{th} response category,

E_i = expected frequency for the i^{th} response category, and

n_r = number of response categories.

Both of these statistics are asymptotically distributed as chi-square (given $n_r \geq k+1$) with degrees of freedom

$$df_r = n_r - k - 1 \qquad (3)$$

where k is the number of independent parameters defining the model in question.

Note that the observed and expected frequencies on which these statistics are based may correspond to either "total scores" or item "response patterns." The preferability of these two classes of response categories will depend on such factors as number of respondents, N, number of items, m, and number of parameters, k, to be estimated.

In addition to assessing the absolute fit of models to data, researchers are frequently interested in comparing the fits provided by different models. When comparisons between pairs of models are of interest, two situations may arise: (1) the models have a subsuming relation, or (2) the models have a non-subsuming relation.

A subsuming relation exists between two models when the parameters of one model (the reduced model) form a subset of those contained in the second more complex model (the subsuming or full model). When models have a subsuming relation, it is possible to assess relative fit of the reduced model to the full model (i.e., to determine whether the reduced model fits as well as the full model). This may be accomplished by using the difference in the likelihood ratio statistics defined in (2) for the full (I) and the reduced (II) models,

$$\chi^2_D = \chi^2_{LR(II)} - \chi^2_{LR(I)} \qquad (4)$$

The difference statistic, χ^2_D, is asymptotically distributed as chi-square (given that the full model holds) with degrees of freedom

$$df_D = k_I - k_{II} \qquad (5)$$

where k_I and k_{II} are respectively the number of independent parameters in models I and II.

Two models have a non-subsuming relation when the defining parameters in neither model form a subset of those defining the other model. A problem arises if a researcher is interested in comparing two models that have a non-subsuming relation, since the difference statistic defined in (4) is not distributed as chi-square. Thus, alternative procedures are needed for assessing the relative adequacy of two models with respect to fit.

The problem of comparing non-subsuming models falls within the context of discriminating between models from separate families. A theory of hypothesis testing for separate families has been developed by Cox (1961, 1962) and extended by Atkinson (1970) whose general approach we follow.

The purpose of this paper is to introduce a general "mixture" probability model which incorporates two non-subsuming probability models and to present a strategy for assessing fit provided by each of the component models.

The Mixture Model

The general model, M_x, presented in this paper incorporates two alternative non-subsuming models (I and II) as weighted components of the full model. Under the Mixture model, the probability of the occurrence of the i^{th} response category is defined as

$$\psi_i = \frac{\theta_i^{1-\lambda} \phi_i^{\lambda}}{\sum_{n_r} \theta_i^{1-\lambda} \phi_i^{\lambda}} \tag{6}$$

In this equation, θ_i and $\emptyset_i$ respectively represent probabilities for the i^{th} response category under the non-subsuming models I and II. λ represents a "mixture" parameter which designates the relative contribution which each component model makes to the probability, Ψ_i. Notice that when λ is set at zero or one, the full model reduces to model I or model II respectively.

Estimation and Fit of the Mixture Model

To obtain maximum likelihood estimates of the parameters which define the component latent structure models, iterative methods such as the Newton-Raphson procedure (see Bock and Lieberman, 1970) or the iterative proportional fitting method(see Clogg, 1977) are frequently employed. Similarly, a conditional maximum likelihood estimate of λ may also be obtained through the use of iterative procedures (see Conte and deBoor, 1972).

A test for the absolute fit of the Mixture model, M_X, can be performed by means of a likelihood ratio statistic, χ^2_{LR}, as defined in (2), where $E_i = N\Psi_i$.[4] This statistic is distributed asymptotically as chi-square with degrees of freedom equal to $n_r - k_I - k_{II} - 2$ (given that estimates of the parameters defining Ψ_i are independent).

Strategy for Assessing Fit

In this section a stagewise strategy for selecting a preferred model is presented. Under this strategy it is possible to reach any one of the following conclusions related to model preference based on fit:

(a) model I provides acceptable fit, and is preferred,

(b) model II provides acceptable fit, and is preferred,

(c) both models provide acceptable fit, but neither is preferred, or

(d) both models provide unacceptable fit.

A detailed description of the strategy for selecting a preferred model follows, with the flow diagram in Figure 1 reflecting this strategy.

I. Assess the absolute fit of the Mixture model, M_X, by means of the χ^2_{LR} statistic in equation (2).

   A. If absolute fit is obtained, go to II.

   B. If absolute fit is not obtained, STOP. Conclude that neither Model I nor Model II (separately or combined) are acceptable for describing the data.

II. (From I.A) Assess the relative fit of Model I and Model II to the Mixture model by means of the difference chi-square, χ^2_D, as defined in (4).

   A. If both Model I and Model II provide acceptable fit, go to III.

   B. If only Model I provides acceptable relative fit, go to IV.

   C. If only Model I provides acceptable relative fit, go to V.

   D. If neither Model I nor Model II provide acceptable relative fit, STOP. Conclude that neither Model I nor Model II are adequate to account for the data.

III. (From II.A) Assess Model I and Model II with respect to their absolute fits with equation (2).

   A. If only Model II provides acceptable fit, STOP. Select Model II.

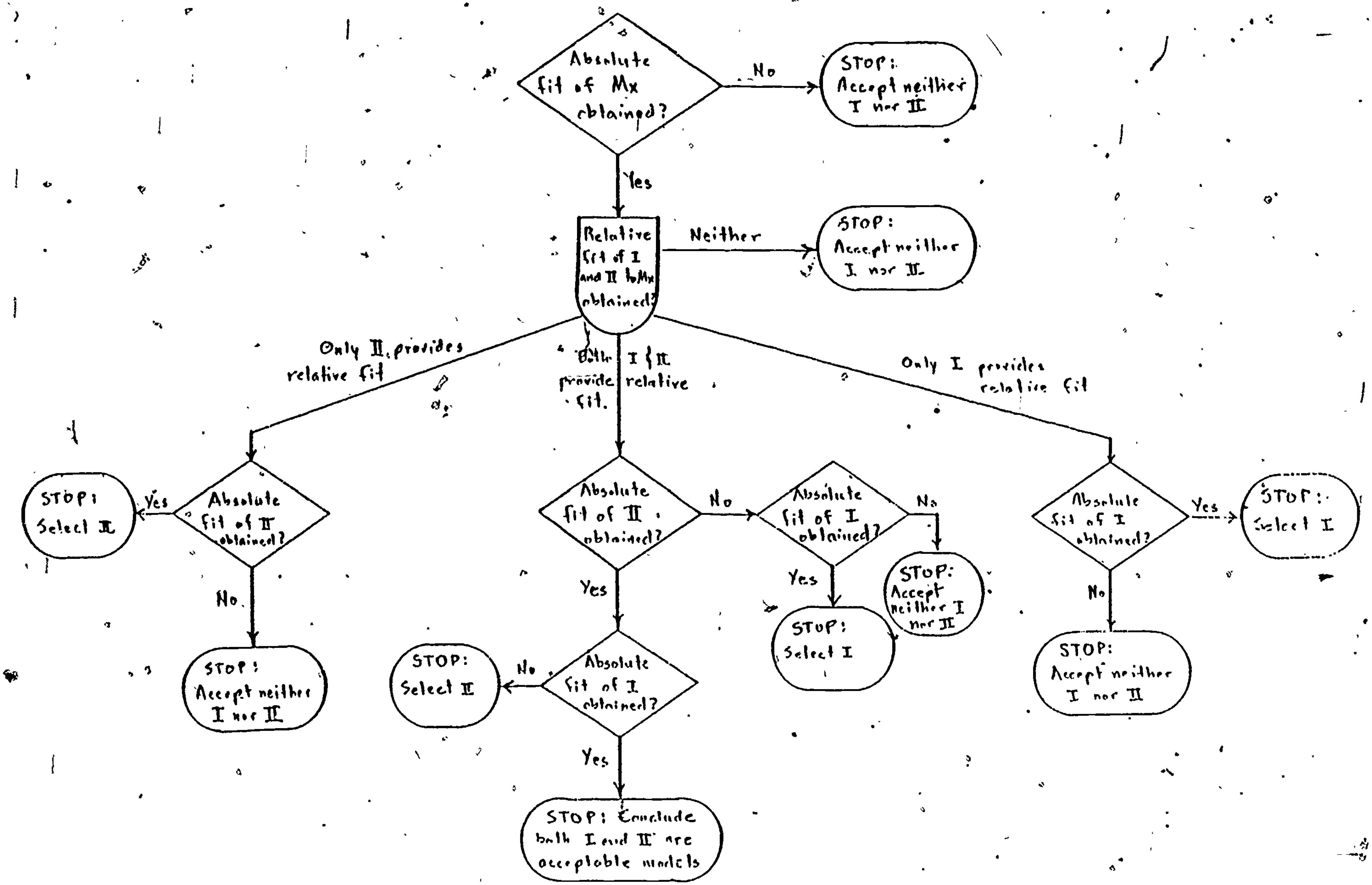

Figure 1. Flow diagram of the proposed strategy for selecting a preferred model.

B. If only Model I provides acceptable fit, STOP. Select Model I.

C. If both Model I and Model II provide acceptable fit, STOP. Conclude that both Model I and Model II are acceptable models but neither is "preferred" over the other model.

D. If neither model provides acceptable absolute fit, STOP. Conclude that neither model is acceptable.

IV. (From II.B) Assess the absolute fit of Model II with equation (2).

A. If absolute fit is obtained, STOP. Select Model II.

B. If absolute fit is not obtained, STOP. Conclude that neither model is acceptable.

V. (From II.C) Assess the absolute fit of Model I with equation (2).

A. If absolute fit is obtained, STOP. Select Model I.

B. If absolute fit is not obtained, STOP. Conclude that neither model is acceptable.

Example Applications

In the area of mastery assessment two classes of latent structure models have been proposed. These have been called Continuum models and State models (see Meskauskas, 1976). For Continuum models, trait acquisition is assumed to be gradual and mastery is viewed as an interval on a continuum while for State models, trait acquisition is conceived as an "all-or-none" process and mastery is viewed as the presence of trait acquisition. Two models which fall respectively within the Continuum and State model classes (and which do not have a subsuming relation) are the Rasch model (Wright and Stone, 1979) and the Latent State Mastery

(LSM) model (Macready and Dayton, 1977). This section of the paper is devoted to the application of the proposed strategy with simulated data, for selection between the Rasch and LSM models.

For both the Rasch and LSM models, the probability of the occurrence of the i^{th} response pattern may be designated as the product of item response probabilities (assuming local independence).

$$P(\underline{u}_i) = \prod_{g=1}^{m} P_g^{\,h_{gi}} (1-P_g)^{1-h_{gi}} \qquad (7)$$

where $\underline{u}_i$ = the i^{th} response pattern

m = the number of dichotomously scored items, and

$$h_{gi} = \begin{cases} 1 & \text{when a correct response is encountered for the } g^{th} \text{ item within the } i^{th} \text{ response pattern} \\ 0 & \text{otherwise} \end{cases}$$

For the Rasch model, the probability of a correct response to item g over all persons is

$$\hat{P}_g = \frac{1}{N} \sum_{j=1}^{N} \frac{\exp\left[\bar{a}\,(x_j - b_g)\right]}{1 + \exp\left[\bar{a}\left(x_j - b_g\right)\right]} \qquad (8)$$

where

x_j = the latent ability of the j^{th} examinee,

b_g = difficulty of the g^{th} item, and

$\bar{a}$ = common discrimination factor for all items.

Correspondingly, for the LSM model, the probability, P_g, is defined as

$$P_g = (P_{g.M})\,(\Delta) + (P_{g.\bar{M}})\,(1 - \Delta) \qquad (9)$$

where

$P_{g.M}$ = the conditional probability that a master (M) answers item g correctly,

$P_{g.\overline{M}}$ = the conditional probability that a non-master ($\overline{M}$) answers item g correctly, and

Δ = the latent proportion of masters (thus, $1 - \Delta$ is the latent proportion of non-masters).

In equation (6), $P(\underline{u}_j)$ is represented as $\cancel{\theta_j}$ for the Rasch model and ϕ_j for the LSM model.

Data Generation

Two sets of simulated data were considered in this paper. The first set of data was based onthe Rasch model for a sample of N = 100 simulated subjects responding to m = 5 items. The item difficulties used in generation were uniformly distributed (-.9, -.45, 0.0, .45, .9) and the discrimination factor ($\overline{a}$) was set equal to .85. The latent trait, x_j, was randomly generated from a normal population with a mean of 0.0 and standard deviation of 1.0.

The second set of simulated data was based on the LSM model, here also with a sample of N=100 simulated subjects responding to m = 5 items. Under this model, the five conditional probabilities for positive item responses from non-masters, $P_{g.\overline{M}}$, were set at .99, .98, .97, .96, and .95.

For each set of simulated data, maximum likelihood estimates of the parameters defining each model were independently estimated through the use of separate programs employing Newton-Raphson algorithms (see Wright, Mead, and Bell, 1979; and Dayton and Macready, 1976). Conditional upon these estimates, a maximum likelihood estimate of λ was obtained using the method of bisection (Conte and deBoor, 1972). In addition, models were judged as providing acceptable absolute (or relative) fit if $(P(\chi^2) \geq .05$.

The results related to the data generated from the Rasch model are presented in Table 1.

Table 1
Results for tests of fit based on data generated from the Rasch model

Assessed Condition	χ^2	df	p-value
Absolute fit of M_x	22.42	14	.930
Relative fit of Rasch	7.17	12	.154
Relative fit of LSM	9.34	6	.845
Absolute fit of Rasch	29.59	26	.715
Absolute fit of LSM	31.76	20	.954

Following the detailed strategy for selecting a preferred model, the absolute fit of the Mixture model, M_x, was first assessed (at Stage I). Since the observed χ^2_{LR} for the Mixture model suggested adequate fit, the relative fits for both the Rasch and LSM models to the mixture model were considered (at Stage II). Because each of these models provided acceptable relative fit to the Mixture model, they were then assessed with respect to their absolute fit (at Stage III). Only the χ^2_{LR} for the Rasch model failed to exceed the .05 critical value. Hence, the Rasch model was selected as the preferred model.

Corresponding results for the data set that was generated from the LSM model are presented in Table 2.

Table 2

Results for tests of fit based on data generated from the LSM model

Assessed Condition	χ^2	df	p-value
Absolute fit of M_x	17.19	14	.754
Relative fit of Rasch	22.03	12	.963
Relative fit of LSM	1.41	6	.035
Absolute fit of Rasch	39.22	26	.954
Absolute fit of LSM	18.61	20	.453

As in the previous example, the Mixture model provided acceptable absolute fit (at Stage I). Thus, (at Stage II) both the Rasch and LSM models were assessed with respect to their relative fits to the Mixture model. Here, only the LSM model provided acceptable relative fit. For this reason only the LSM model was assessed with respect to its absolute fit (at Stage IV) which was obtained. Consequently, the LSM model was selected as the preferred model.

In each of the examples considered it may be seen that the decision strategy led to the selection of the generating model.

References

Atkinson, A. C., "A method for discriminating between models," J. Roy. Stat. Soc., 3, 1970, 323-345.

Bock, R. D., and Lieberman, M., "Fitting a response model for n dichotomously scored item," Psychometrika, 35, (2), 1970, 179-197.

Clogg, C. C., "Unrestricted and restricted maximum likelihood latent structure analysis: A manual for users," Working Paper #1988-09, 1977.

Conte, S. D., and deBoor, C., Elementary Numerical Analysis, McGraw-Hill, New York, 1972.

Cox, D. R., "Tests on separate families of hypotheses," Proceedings of the 4th Berkely Symposium, 1, 1961, 105-123.

Cox, D. R., "Further results on tests of separate families of hypotheses," J. Roy. Stat. Soc., B, 24, 1962, 406-424.

Dayton, C. M. and Macready, G. B., "A propabilistic model for validation of behavioral hierarchies," Psychometrika, 1976, 41, 189-204.

Meskauskas, J. A., "Evaluation models for criterion referenced testing: views regarding mastery and standard setting," Review of Educational Research, 1976, 46, 133-158.

Macready, G. B., and Dayton, C. M., "The use of probabilistic models in the assessment of mastery," Journal of Educational Statistics, 1977, 2, (2), 99-120.

Wright, B. D., Mead, R. J., and Bell, S., "BICAL" Computer program, Statistical Laboratory, Department of Education, University of Chicago, 1976.

Wright, B. D., and Stone, M. H., Best Test Design, MESA, 1979.